THE RELATION BETWEEN MEASURES OF CORRELATION IN THE UNIVERSE OF SAMPLE PERMUTATIONS

By H. E. DANIELS, *Wool Industries Research Association*

1. Introduction

Recent papers by Hotelling & Pabst (1936), Pitman (1937), Kendall (1938) and Kendall, Kendall & Babington-Smith (1939) discuss the distribution of the correlation coefficient when members of the sample corresponding to the two variates are permuted randomly relative to one another. In the case of rank correlation, the characteristics of the population sampled are generally unknown, and a significance test has to be based on the distribution obtained from the sample in this way.

Hotelling & Pabst prove that as the sample size is increased, Spearman's ρ tends to follow a normal distribution law. Kendall's measure of rank correlation, τ, in which all possible corresponding pairs in two given rankings are assigned marks ± 1 according to whether they agree or differ in order, follows a specially simple distribution law which tends rapidly to the normal form and becomes highly correlated with Spearman's ρ for samples of moderate size.

The present paper discusses the properties of the class of correlation coefficients Γ obtained on replacing Kendall's marks ± 1 by a more general system of scores. By an empirical argument Kendall *et al.* showed it to be likely that the correlation between τ and ρ is $\dfrac{2(n+1)}{\sqrt{[2n(2n+5)]}}$ for all values of the sample size n, and surmised that their joint distribution tends to the bivariate normal form for large n. These results are, in fact, special cases of the relations demonstrated below between two correlation coefficients Γ with different systems of scores.

2. Definition

Consider the two sets of n sample values

$$x_1, x_2, \ldots, x_n, \quad y_1, y_2, \ldots, y_n,$$

both arranged in some given order relative to each other. They may be permuted to give $n!$ different ways of grouping the x's with the y's. Let us assign to each pair (x_i, x_j) what for convenience will be termed a score a_{ij} and to each (y_i, y_j) a score b_{ij}, where

$$a_{ij} = -a_{ji}, \quad b_{ij} = -b_{ji}.$$

Denote by Γ the number
$$\Gamma = \frac{\Sigma a_{ij} b_{ij}}{\sqrt{(\Sigma a_{ij}^2\, \Sigma b_{ij}^2)}},$$

the summation extending over all i and j from 1 to n. Special cases of Γ are Kendall's τ, the product-moment correlation coefficient r and Spearman's ρ, for τ is obtained by definition when $a_{ij}, b_{ij} = \pm 1, j \gtrless i$, r is given when $a_{ij} = x_j - x_i$, $b_{ij} = y_j - y_i$ by virtue of the identity

$$\tfrac{1}{2} \sum_i \sum_j (x_j - x_i)(y_j - y_i) \equiv n \sum_i x_i y_i - \sum_i \sum_j x_i y_j,$$

and ρ is similarly obtained when $a_{ij}, b_{ij} = j - i$.

When the x's are permuted relative to the y's, the scores reappear in a new order with the same or opposite sign and the denominator of Γ remains unaltered, so that in discussing the distribution of Γ over all permutations it is sufficient to consider the numerator only, which we denote by c.

Write A, B for the matrices of the scores; for example, with $n = 4$,

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{bmatrix}; \qquad B = \begin{bmatrix} 0 & b_{12} & b_{13} & b_{14} \\ -b_{12} & 0 & b_{23} & b_{24} \\ -b_{13} & -b_{23} & 0 & b_{34} \\ -b_{14} & -b_{24} & -b_{34} & 0 \end{bmatrix}.$$

With the x's and y's in the order as written, c is the trace of the matrix product AB' (i.e. the sum of the elements of its leading diagonal), where B' is the transpose of B. The effect of a permutation of the x's, say, is to alter the score matrix of the x pairs to PAP' and the value of c to the trace of $PAP'B'$, where $P = (p_{ij})$ is the appropriate 'permutation matrix' obtained by permuting the columns of the unit matrix. For example, corresponding to the grouping

$$x_4, x_1, x_2, x_3, \quad y_1, y_2, y_3, y_4,$$

the permutation matrix is

$$P = \begin{bmatrix} . & . & . & 1 \\ 1 & . & . & . \\ . & 1 & . & . \\ . & . & 1 & . \end{bmatrix}.$$

In terms of the matrix elements c is given by

$$c = \Sigma p_{ji} p_{kl} a_{il} b_{jk},$$

all suffixes being summed from 1 to n.

3. Moments

The distribution of c over all permutations, or the joint distribution of two c's with different systems of scores, is most readily discussed from its moments. The product moment of $c(1)$ and $c(2)$ with scores $a^{(1)}_{ij}$, $b^{(1)}_{ij}$ and $a^{(2)}_{ij}$, $b^{(2)}_{ij}$ respectively is

$$\overline{c(1)\,c(2)} = \frac{1}{n!} S(c(1)\,c(2)),$$

where

$$c(1)\,c(2) = \Sigma p_{ji} p_{kl} p_{sr} p_{ut} a^{(1)}_{il} b^{(1)}_{jk} a^{(2)}_{rt} b^{(2)}_{su},$$

and S denotes summation over all $n!$ possible permutations. Consider the effect of S on each term. The non-vanishing contributions occur when all four p's are 1, and it is first noted that if any of the suffixes i, l, r, t are equal, the corresponding suffixes in j, k, s, u must also be equal for the term not to vanish, since each row of P contains only one non-vanishing element. Terms in which $i = l$ or $r = t$ are of course zero by definition. When, for example, $i = r$ so that r is replaced by i in the expression, we shall call i a *tied suffix*. Other suffixes will be referred to as *free* suffixes.

As regards their contribution to S the terms may be classified according to the number of tied suffixes in i, l, r, t as follows.

(i) *No tied suffixes*. When the four p's are each unity, four rows and columns of P are assigned and there are $(n-4)!$ ways of filling the remaining positions. Such terms therefore contribute

$$(n-4)!\, \Sigma' a^{(1)}_{il} a^{(1)}_{rt} \Sigma' b^{(1)}_{jk} b^{(2)}_{su}$$

to the total sum, where Σ' denotes summation over all values of the suffixes which are not equal. Let us consider some properties of $\Sigma' a_{il} a_{rt}$ and $\Sigma a_{il} a_{rt}$ and similar expressions with tied suffixes, the second expression being summed over *all* values of the suffixes. (Superscripts (1) and (2) are understood throughout.)

(1) $\Sigma a_{ii} a_{rt} = 0$, $\Sigma' a_{ii} a_{rt} = 0$, and so on.

(2) $\Sigma a_{il} a_{il} = \Sigma' a_{il} a_{il}$.

(3) $\Sigma a_{il} a_{it} = \Sigma' a_{il} a_{it} + \Sigma' a_{il} a_{il}$.

(4) $\Sigma a_{il} a_{rt} = 0$, $\Sigma' a_{il} a_{rt} = 0$. The first is true because $\Sigma a_{il} = 0$, and the second follows from the fact that

$$\Sigma a_{il} a_{rt} = \Sigma' a_{il} a_{rt} + \Sigma' a_{il} a_{it} + \Sigma' a_{il} a_{ri} + \Sigma' a_{il} a_{ll} + \Sigma' a_{il} a_{rl} + \Sigma' a_{il} a_{il} + \Sigma' a_{il} a_{li},$$

the terms on the right after the first cancelling in pairs.

The contribution to S of terms with no tied suffixes is therefore zero.

(ii) *One tied suffix*. For the term not to vanish it is necessary to assign three rows and columns of P, and the contribution to S from such terms is

$$4(n-3)!\,\Sigma' a_{il}^{(1)} a_{it}^{(2)} \Sigma' b_{jk}^{(1)} b_{ju}^{(2)},$$

the factor 4 arising from the fact that the same contribution is obtained by tying the suffixes in the four possible ways.

(iii) *Two tied suffixes*. The contribution to S is similarly found to be

$$2(n-2)!\,\Sigma' a_{il}^{(1)} a_{il}^{(2)} \Sigma' b_{jk}^{(1)} b_{jk}^{(2)}.$$

Terms containing more than two tied suffixes give zero contributions to S, and finally, substituting for Σ' the appropriate Σ expressions, we find

$$\overline{c(1)\,c(2)} = \frac{4}{n(n-1)(n-2)} (\Sigma a_{il}^{(1)} a_{it}^{(2)} - \Sigma a_{il}^{(1)} a_{il}^{(2)}) (\Sigma b_{jk}^{(1)} b_{ju}^{(2)} - \Sigma b_{jk}^{(1)} b_{jk}^{(2)}) + \frac{2}{n(n-1)} \Sigma a_{il}^{(1)} a_{il}^{(2)} \Sigma b_{jk}^{(1)} b_{jk}^{(2)}.$$

The moments of higher order can be obtained by a similar procedure, but the expressions rapidly become unwieldy.

4. THE CORRELATION BETWEEN KENDALL'S τ AND SPEARMAN'S ρ

As a first application of the formula we consider the correlation between τ and ρ over all permutations of the sample values. The scores for τ and ρ respectively are

$$a_{ij}^{(1)},\ b_{ij}^{(1)} = \pm 1,\ 0 \quad \text{when} \quad j \gtrless i,\ j = i,$$

$$a_{ij}^{(2)},\ b_{ij}^{(2)} = j - i.$$

The following results are easily derived

$$\sum_{l=1}^{n} a_{il}^{(1)} = n + 1 - 2i, \quad \sum_{l=1}^{n} a_{il}^{(2)} = \tfrac{1}{2}n(n+1-2i),$$

$$\sum_{i=1}^{n} \sum_{l=1}^{n} \sum_{t=1}^{n} a_{il}^{(1)} a_{it}^{(2)} = \frac{n^2(n^2-1)}{6}, \quad \sum_{i=1}^{n} \sum_{l=1}^{n} a_{il}^{(1)} a_{il}^{(2)} = \frac{n(n^2-1)}{3},$$

and the same results hold for the b's. Substitution in the formula then gives

$$\overline{c(1)\,c(2)} = \frac{n^2(n-1)(n+1)^2}{9}.$$

Again, $$\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n} a_{il}^{(1)}a_{it}^{(1)} = \frac{n(n^2-1)}{3}, \quad \sum_{i=1}^{n}\sum_{l=1}^{n} a_{il}^{(1)}a_{il}^{(1)} = n(n-1)$$

and $$\sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n} a_{il}^{(2)}a_{it}^{(2)} = \frac{n^3(n^2-1)}{12}, \quad \sum_{i=1}^{n}\sum_{l=1}^{n} a_{il}^{(2)}a_{il}^{(2)} = \frac{n(n^2-1)}{3},$$

from which it is found that

$$\overline{c(1)^2} = \frac{2n(n-1)(2n+5)}{9}, \quad \overline{c(2)^2} = \frac{n^4(n-1)(n+1)^2}{36}.$$

The required correlation is therefore

$$R_{\rho\tau} = \frac{2(n+1)}{\sqrt{[2n(2n+5)]}}$$

which is the result anticipated by Kendall *et al.* It should be noted that they use the quantities $\Sigma = \frac{1}{2}c(1)$ and

$$S(d^2) = \frac{n(n^2-1)}{6} - \frac{c(2)}{n}$$

in place of $c(1)$ and $c(2)$.

5. Transformation of the sample values

If the scales of the x's and y's are distorted by a transformation and the product moment correlation coefficient r is recalculated on the transformed sample, a new value r' is obtained. In particular, the x's and y's may be readjusted to be at equidistant intervals, and then the new value of r is Spearman's ρ. The formula for $\overline{c(1)\,c(2)}$ can be used to find the correlation over all sample permutations between the values of r on the same sample before and after such a transformation. Distinguishing by primes the sample values after transformation, the scores are

$$a_{ij}^{(1)} = x_j - x_i, \quad b_{ij}^{(1)} = y_j - y_i,$$

$$a_{ij}^{(2)} = x'_j - x'_i, \quad b_{ij}^{(2)} = y'_j - y'_i.$$

Then $$\sum_{l=1}^{n} a_{il}^{(1)} = n(\bar{x} - x_i), \quad \sum_{i=1}^{n}\sum_{l=1}^{n}\sum_{t=1}^{n} a_{il}^{(1)}a_{it}^{(2)} = n^2\sum_{i=1}^{n}(x_i-\bar{x})(x'_i-\bar{x}'),$$

$$\sum_{i=1}^{n}\sum_{l=1}^{n} a_{il}^{(1)}a_{il}^{(2)} = \sum_{i=1}^{n}\sum_{l=1}^{n}(x_l-x_i)(x'_l-x'_i) = 2n\Sigma(x_i-\bar{x})(x'_i-\bar{x}')$$

by the identity previously quoted. Using these and similar formulae we find

$$\overline{c(1)\,c(2)} = \frac{4n^2}{n-1}\Sigma(x-\bar{x})(x'-\bar{x}')\Sigma(y-\bar{y})(y'-\bar{y}'),$$

$$\overline{c(1)^2} = \frac{4n^2}{n-1}\Sigma(x-\bar{x})^2\Sigma(y-\bar{y})^2; \quad \overline{c(2)^2} = \frac{4n^2}{n-1}\Sigma(x'-\bar{x}')^2\Sigma(y'-\bar{y}')^2,$$

and hence the correlation between r and r' is

$$R_{rr'} = r_{xx'}r_{yy'},$$

where $r_{xx'}$ and $r_{yy'}$ are the correlation coefficients between old and new values of x and y respectively.

6. Tendency to normal form for large n

It will now be shown for a large class of score systems a_{ij} that c, and hence Γ, tends with increasing n to be normally distributed, and moreover, that the joint distribution of any pair of such Γ's tends to the bivariate normal form.

The pth order product moments of the joint distribution of $c(1)$ and $c(2)$ are sums of terms containing

$$\Sigma' a_{gh} a_{ij} a_{kl} \ldots \Sigma' b_{rs} b_{tu} b_{vw} \ldots$$

or similar expressions in which arbitrary groups of suffixes within the Σ''s are tied, each Σ' involving products of p scores which may belong either to systems (1) or (2). Every such Σ' is in turn a linear combination of the corresponding Σ having the same suffixes and other Σ's in which additional tied suffixes are introduced. No Σ may contain a pair of free suffixes attached to one score, for it would then vanish by virtue of the fact that $\Sigma a_{ij} = 0$.

The even order product moments are first discussed. Let $p = 2m$. Consider a Σ in which the $2m$ scores are divided into m pairs each having one tied suffix, so that there are in all $3m$ independent suffixes, e.g.

$$\Sigma a_{ij} a_{ik} a_{lr} a_{ls} a_{tu} a_{tv} \ldots.$$

It may be written as

$$(\Sigma a^{(1)}_{ij} a^{(1)}_{ik})^{\lambda} \, (\Sigma a^{(1)}_{ij} a^{(2)}_{ik})^{\mu} \, (\Sigma a^{(2)}_{ij} a^{(2)}_{ik})^{\nu},$$

where $\lambda + \mu + \nu = m$ and λ, μ, ν are the number of times the scores are paired in the combinations indicated.

As is always possible, suppose the numerically largest value of a_{ij} to be made equal to unity. We now impose the condition that $\Sigma a_{ij} a_{ik}$ is of order n^3 whether a_{ij} and a_{ik} belong to the same or different systems of scores. This is satisfied, when $\max a_{ij} = 1$, by τ and ρ, and also by r provided the sample is not an unusual one. With this condition, it is seen that Σ's of the above type are of order n^{3m}.

It is next observed that all other ways of tying suffixes give Σ's of lower order of magnitude. For the order of magnitude of the bracket is not reduced on replacing each a_{ij} by $+1$; consequently if further suffixes are tied the order of Σ is made less than n^{3m} since there are fewer than $3m$ summations from 1 to n. It follows that the dominant term in a Σ' is the corresponding Σ having the same array of suffixes.

Moreover, every non-vanishing Σ involving $3m$ independent suffixes can only be a permutation of the type illustrated, while those with more than $3m$ different suffixes must all vanish. This is made clear by considering how the $3m$ suffixes can be arrayed between the $2m$ scores. Begin by assigning $3m$ different suffixes at random among the $4m$ available places. At least m scores will receive their full complement of suffixes all which will be different. There cannot be more than m such completed scores, for if Σ is not to vanish, at least one suffix of each complete pair must be tied and this can only be done by repeating one suffix from every complete pair in each of the remaining places to be filled, of which there are only m. We are thus led to a permutation of the type of Σ discussed above. If there had been more than $3m$ different suffixes to begin with, there would not have remained sufficient empty places to prevent the existence of at least one score with a pair of free suffixes, and so all Σ's with more than $3m$ different suffixes must vanish.

Any $2m$th product moment is the sum of terms like

$$\frac{(n-f)!}{n!} A \Sigma' a_{ij} a_{kl} \ldots \Sigma' b_{rs} b_{tu} \ldots,$$

where f is the number of independent suffixes in the Σ''s and A is a coefficient which is of unit order as far as n is concerned. From the preceding argument, the maximum value of f is $3m$, in which case the term is of order $n^{-3m} \times n^{3m} \times n^{3m} = n^{3m}$. When $f \leqslant 3m-1$ the order of the term is not greater than $n^{-3m+1} \times n^{3m-1} \times n^{3m-1} = n^{3m-1}$ and such terms may therefore be neglected. Write

$$h_{11} = \Sigma a_{ij}^{(1)} a_{ik}^{(1)} \Sigma b_{tu}^{(1)} b_{tv}^{(1)}, \quad h_{12} = \Sigma a_{ij}^{(1)} a_{ik}^{(2)} \Sigma b_{tu}^{(1)} b_{tv}^{(2)}, \quad h_{22} = \Sigma a_{ij}^{(2)} a_{ik}^{(2)} \Sigma b_{tu}^{(2)} b_{tv}^{(2)}.$$

Then if terms of lower order of magnitude are neglected, the even product moment

$$\mu_{r,s} = \overline{c(1)^r c(2)^s}, \quad r+s = 2m$$

is given by the sum of terms like

$$n^{-3m} A_{\lambda,\mu,\nu} h_{11}^{\lambda} h_{12}^{\mu} h_{22}^{\nu}; \quad 2\lambda + \mu = r, \quad \mu + 2\nu = s,$$

over all possible values of λ, μ, ν. The coefficient $A_{\lambda,\mu,\nu}$, which is the number of ways in which $h_{11}^{\lambda} h_{12}^{\mu} h_{22}^{\nu}$ can arise, is calculated as follows. Consider a Σ whose array of suffixes is such that it can be factorized as $(\Sigma a_{ij}^{(1)} a_{ik}^{(1)})^{\lambda} (\Sigma a_{ij}^{(1)} a_{ik}^{(2)})^{\mu} (\Sigma a_{ij}^{(2)} a_{ik}^{(2)})^{\nu}$. Its suffix pairs can be permuted within the sets of scores (1) and (2) in $r!\,s!$ ways, but of these $\lambda!(2!)^{\lambda}\,\mu!\,\nu!(2!)^{\nu}$ give essentially the same Σ. The suffixes within pairs attached to each score may also be rearranged in 2^{2m} ways without affecting the result, and so

$$A_{\lambda,\mu,\nu} = \frac{r!\,s!\,2^{2m}}{\lambda!\,\mu!\,\nu!\,2^{\lambda+\nu}} = \frac{r!\,s!\,2^{m+\mu}}{\lambda!\,\mu!\,\nu!}.$$

The calculation of the even order product moment $\mu_{r,s}$ for large n is in fact tantamount to selecting the coefficient of $t^r t^s / r!\,s!$ in

$$\frac{2^m}{n^{3m} m!} (h_{11} t_1^2 + 2h_{12} t_1 t_2 + h_{22} t_2^2)^m.$$

Finally, we dispose of the odd moments. In certain cases, such as for example the joint distribution of τ and ρ, they all vanish by symmetry. But even in the general case it can be shown that the odd moments are negligible to the order of magnitude $n^{-\frac{1}{2}}$.

A Σ containing $2n+1$ scores cannot have more than $3m+1$ different suffixes. For if there were $3m+2$, let them first be assigned to the $4m+2$ available places; at least $m+1$ scores will receive complete pairs of suffixes, and the remaining m empty places cannot be filled in any way which avoids one score having a free pair of suffixes. Hence as before the order of magnitude of any $(2m+1)$th moment is at most $n^{-(3m+1)} \times n^{3m+1} \times n^{3m+1} = n^{3m+1}$.

The $2m$th moments were shown to be of order n^{3m}, consequently if we define

$$\gamma(1) = n^{-\frac{3}{2}} c(1), \quad \gamma(2) = n^{-\frac{3}{2}} c(2),$$

the joint distribution of $\gamma(1)$ and $\gamma(2)$ has all its even moments of unit order, and by the result just proved all its odd moments are of order $n^{-\frac{1}{2}}$ and may therefore be neglected to that order. Reverting to $c(1)$, $c(2)$, it is seen that the moment-generating function of their joint distribution tends in the limit to the form

$$\exp \frac{2}{n^3} (h_{11} t_1^2 + 2h_{12} t_1 t_2 + h_{22} t_2^2).$$

Hence $c(1)$ and $c(2)$ tend to be normally distributed with variances $\frac{4}{n^3} h_{11}$, $\frac{4}{n^3} h_{22}$ and correlation

$$\frac{h_{12}}{\sqrt{(h_{11} h_{22})}} = \frac{\Sigma a_{ij}^{(1)} a_{ik}^{(2)}}{\sqrt{(\Sigma a_{ij}^{(1)} a_{ik}^{(1)} \Sigma a_{ij}^{(2)} a_{ik}^{(2)})}} \frac{\Sigma b_{tu}^{(1)} b_{tv}^{(2)}}{\sqrt{(\Sigma b_{tu}^{(1)} b_{tv}^{(1)} \Sigma b_{tu}^{(2)} b_{tv}^{(2)})}}.$$

The Γ's similarly tend to a bivariate normal distribution with the same correlation, but with variances,

$$\frac{4}{n^3}\frac{\Sigma a^{(1)}_{ij}a^{(1)}_{ik}\,\Sigma b^{(1)}_{tu}b^{(1)}_{tv}}{\Sigma a^{(1)}_{ij}a^{(1)}_{ij}\,\Sigma b^{(1)}_{tu}b^{(1)}_{tu}},\quad \frac{4}{n^3}\frac{\Sigma a^{(2)}_{ij}a^{(2)}_{ik}\,\Sigma b^{(2)}_{tu}b^{(2)}_{tv}}{\Sigma a^{(2)}_{ij}a^{(2)}_{ij}\,\Sigma b^{(2)}_{tu}b^{(2)}_{tu}}.$$

Our proof rests on the assumption that $\Sigma a_{ij}a_{ik}$ and $\Sigma b_{tu}b_{tv}$ are of order n^3, where the individual a_{ij}'s and b_{tu}'s may belong to either score system. But if that is true, it follows that expressions like Σa^2_{ij} must be of order n^2, for they cannot be made to exceed that order on replacing a_{ij} by $+1$, and their order cannot be less than n^2 since

$$\Sigma a^2_{ij} - \frac{1}{n}\Sigma a_{ij}a_{ik} = \Sigma(a_{ij}-\bar{a}_i)^2 \geqslant 0,$$

where $\bar{a}_i = \frac{1}{n}\sum_{j=1}^{n} a_{ij}$. Consequently the variances of the Γ's decrease like n^{-1}. The correlation between the Γ's tends, however, to a value independent of n in the limit.

SUMMARY

The properties of a general class of correlation coefficients Γ, which includes the product-moment correlation coefficient r, Spearman's ρ and Kendall's τ, are discussed. A direct proof is given of the formula tentatively suggested by Kendall for the correlation between ρ and τ when the sample is permuted in all possible ways. The effect of a transformation of the sample values is also considered. It is shown that under certain general conditions, the joint distribution of two different Γ's, calculated on all possible permutations of the sample values, tends with increasing sample size to the bivariate normal form with variances inversely proportional to the sample size and correlation independent of it.

REFERENCES

HOTELLING, H. & PABST, M. R. (1936). Rank correlation and tests of significance involving no assumptions of normality. *Ann. Math. Statist.* **7**, 29.

PITMAN, E. J. G. (1937). Significance tests which may be applied to samples from any populations. II. The correlation coefficient test. *J.R. Statist. Soc. Suppl.* **4**, 225.

KENDALL, M. G. (1938). A new measure of rank correlation. *Biometrika*, **30**, 81.

KENDALL, M. G., KENDALL, S. F. H. & BABINGTON-SMITH, B. (1939). The distribution of Spearman's coefficient of rank correlation in a universe in which all rankings occur an equal number of times. *Biometrika*, **30**, 251.